AFRL-RV-PS-
TR-2016-0096

MACHINE LEARNING AIDED EFFICIENT AND ROBUST ALGORITHMS FOR SPECTRUM KNOWLEDGE ACQUISITION IN WIDEBAND AUTONOMOUS COGNITIVE RADIOS

Sudharman Jayaweera

Department of Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM 87131

1 Aug 2016

Final Report

AIR FORCE RESEARCH LABORATORY
Space Vehicles Directorate
3550 Aberdeen Ave SE
AIR FORCE MATERIEL COMMAND
KIRTLAND AIR FORCE BASE, NM 87117-5776

DTIC COPY
NOTICE AND SIGNATURE PAGE

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RV-PS-TR-2016-0096 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signed//
KHANH PHAM
Program Manager

//signed//
DAVID CARDIMONA
Technical Advisor, Space Based Advanced Sensing and Protection

//signed//
JOHN BEAUCHEMIN
Chief Engineer, Spacecraft Technology Division
Space Vehicles Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-08-2016	Final Report	27 Apr 2015 to 27 Jul 2016

4. TITLE AND SUBTITLE

Machine Learning Aided Efficient and Robust Algorithms for Spectrum Knowledge Acquisition in Wideband Autonomous Cognitive Radios

5a. CONTRACT NUMBER
FA9453-15-1-0314

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
63401F

6. AUTHOR(S)

Sudharman Jayaweera

5d. PROJECT NUMBER
682J

5e. TASK NUMBER
PPM00016000

5f. WORK UNIT NUMBER
EF125354

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Department of Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM 87131

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Space Vehicles Directorate
3550 Aberdeen Ave SE
Kirtland AFB, NM 87117-5776

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RVSW

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-RV-PS-TR-2016-0096

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The objective of this project was to conduct research that will advance the wideband autonomous cognitive radio (WACR) technology. These are radios that have the ability to sense state of the radio frequency (RF) spectrum and the network and self-optimize its operating mode in response to this sensed state. First, a formal framework was developed for robust spectrum knowledge acquisition in a wideband autonomous cognitive radio. The performance of this framework was evaluated based on simulations. Next, a machine learning based sub-band selection algorithm for WACRs was developed based on reinforcement learning and its performance was analyzed through a combination of analysis and simulations. Finally, motivated by certain application scenarios of interest, a new definition for the state of the spectrum of interest to a WACR was developed. Currently, this new approach is being used for developing practical cognitive communications protocols with considerably less computational complexity than previous alternatives. Future work will include design, implementation and analysis of cognitive communications protocols suited for space and satellite communications using this new state definition.

15. SUBJECT TERMS
Space communication; satellite communication; Cognitive radios; machine learning

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Khanh Pham
Unclassified	Unclassified	Unclassified	Unlimited	32	**19b. TELEPHONE NUMBER** *(include area code)*

(This page intentionally left blank)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This material is based on research sponsored by Air Force Research Laboratory under agreement number FA9453-15-1-0314. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

DISCLAIMER

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Air Force Research Laboratory or the U.S. Government.

(This page intentionally left blank)

1 SUMMARY

The objective of this project was to conduct research that will advance the wideband autonomous cognitive radio (WACR) technology. These are radios that have the ability to sense state of the radio frequency (RF) spectrum and the network and self-optimize their operating modes in response to this sensed state. First, this project developed a formal framework for robust spectrum knowledge acquisition in a wideband autonomous cognitive radio. This framework was implemented on a simulation scenario to evaluate its performance. An important functionality in a WACR is the sub-band selection which allows the radio to operate over a wide spectrum range with real-time awareness of the spectrum state. Thus, a machine learning based sub-band selection algorithm for WACRs was developed based on reinforcement learning and its performance was analyzed through a combination of analysis and simulations. Finally, motivated by certain application scenarios of interest, a completely new definition for the state of the spectrum of interest to a WACR was developed. Currently, this new approach is being used for developing practical cognitive communications protocols with considerably less computational complexity than previous alternatives. Future work will include design, implementation and analysis of cognitive communications protocols suited for space and satellite communications using this new state definition.

2 INTRODUCTION

Wideband Autonomous Cognitive Radios (WACRs) proposed in [1] and investigated in this project present a potential future technology to realize autonomous radio communications over non-contiguous wide spectrum bands in the presence of adverse conditions. These adverse conditions may be both deliberate as well as inadvertent. Moreover, encroachment on previously-allocated spectrum resources by commercial and unlicensed/unauthorized users can only be expected to grow in the coming years. Combination of these traditional as well as evolving spectrum demands requires future telecommunications technologies to be intelligent, self-aware and spectrally agile. Wideband autonomous cognitive radios (WACRs) pursued in this project are radios with these defining characteristics that can lead to autonomous radio communications over non-contiguous wide spectrum bands in the presence of adverse conditions [1,2].

Spectrum awareness is the most salient feature of cognitive radios that makes them cognitive and spectrum sensing is the process of acquiring spectrum awareness [1,3,4]. In the case of wideband autonomous cognitive radios, spectrum sensing is usually performed over several non-contiguous spectrum bands each spanning hundreds of MHz to even on the order of a GHz [1,2,5-7]. Due to wide bandwidth and noncontiguous nature of the frequency range of interest, the spectrum knowledge acquisition problem posed by WACR is significantly more challenging than simple spectrum sensing performed by a dynamic spectrum sharing (DSS) cognitive radio. In particular, a wideband cognitive radio is expected to identify all spectral activities present in the spectrum band of interest [1]. These signals can be of different types and be located at unknown carrier frequencies spread over a wide spectrum range. Noise, interference and propagation properties can, however, vary significantly over the spectrum of interest to a wideband autonomous cognitive radio rendering many usual assumptions made in conventional signal processing invalid.

The spectrum knowledge acquisition problem posed by an autonomous wideband cognitive radio can be divided in to three steps as shown on Fig. 1 [1]:

1. Wideband spectrum scanning problem: How to efficiently and effectively scan a wide spectrum range in real-time, or near real-time.
2. Spectral activity detection problem: How to detect the active signals in the sensed spectrum.
3. Signal identification problem: How to classify and identify the origins of the detected signals.

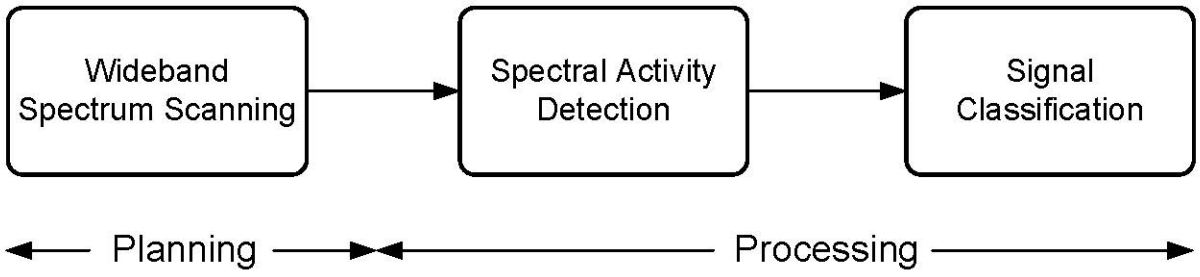


Figure 1. Spectrum knowledge acquisition consists of a planning stage and a processing stage.

In order for a WACR to detect spectrum opportunities it must be able to observe and interpret its surrounding RF environment. The first stage of this process, as shown in Fig. 1, is the wideband spectrum scanning [1,8]. Hardware constraints limit the instantaneous sensing bandwidth of most state-of-the-art software-defined radio (SDR) platforms. Hence, the challenge in this step is to design an efficient scheme to achieve real-time sensing over a wide spectrum range. In order to be able to scan a wide spectrum band in real-time, the spectrum of interest is first divided into a set of sub-bands. Each sub-band can be wide enough to contain multiple communication channels. Since the WACR can sense only one sub-band at any given time, it needs to determine which one to be sensed at each time instant. This problem is known as sub-band selection problem in wideband spectrum sensing [1].

After scanning a spectrum band of interest, the second step is to detect any spectrum activity present in a sensed spectrum sub-band. We must emphasize that the objective of a WACR is to detect and identify all spectral activities present in the spectrum bands of interest to the cognitive radio, not just detecting whitespace in spectrum. Thus, a third step of signal classification and identification may be needed after signal detection [1,7,9].

The long-term objective of this project is to systematically develop a comprehensive wideband spectrum knowledge acquisition framework over the frequencies of interest to satellite and space communications. This will advance the proposed WACR technology leading to future space and satellite radio systems that can be autonomous, self-aware and intelligent.

During the last performance period, we have specifically been focused on the following aspects of this larger project:

- Develop and optimize a compressive sampling-based robust spectrum estimation algorithm suitable for a WACR.
- Developing a machine learning aided sub-band selection algorithm for a WACR based on reinforcement learning paradigm.
- Developing a new mathematical definition of the state of the spectrum specifically suited for the WACR context and using it to design new machine learning aided sub-band selection algorithms.

Continuing from our previous work [10], we have been attempting to optimize our compressive sampling based robust spectrum estimation approach. Note that, compressive

sampling is used to reduce the high sampling rate requirements demanded by wideband spectrum sensing [1,11]. When there is a certain amount of sparsity in the signal with respect to some basis, compressive sampling can be an efficient technique for reconstructing a signal that is sparse with respect to some basis (i.e. most of the expansion coefficients of the signal are zero with respect to a certain basis) [12,13]. The efficiency afforded by compressive sampling is two-fold: First, a smaller number of samples, compared to what is needed with traditional Shannon-Nyquist sampling, will suffice. Second, the reconstruction of the signal from this reduced number of samples can be achieved with an algorithm with low computationally complexity [1].

The suitability of a compressive sampling-based front-end for a cognitive radio hinges on the fact that in many situations sensed sub-bands will have low spectrum utilization making them sparse with respect to a frequency-domain basis. This is indeed usually the case in many spectrum ranges. Even in situations in which a few hundreds of MHz of spectrum may be highly utilized, the frequency-domain sparsity may be applicable based on the instantaneous bandwidth of the spectrum sensing front-end and the chosen sub-band bandwidth [1]. In fact, it is the relatively large sub-band bandwidth that may necessitate a robust spectral activity detector in place of the Gaussian assumption [1]. The robust spectral activity detection approach we have been developing during this project, based on compressive sampling, is aimed at providing robustness against possibly non-Gaussian jammers, interference and other unwanted electromagnetic interference (EMI) [1,10].

The sub-band dynamics model proposed in [1] can reasonably be used to develop effective sub-band selection algorithms. To overcome the unavailability of model parameters and time-varying conditions, machine learning can be incorporated into such decision algorithms. In this project, we have specifically been focused on using reinforcement learning algorithms for this purpose for their suitability in Markov environments [1, 14]. As we will discuss in the next section, however, it may be possible to develop new models for spectrum dynamics based on new mathematical definitions for the spectrum state as needed in particular application scenarios. These new models may lead to learning and decision algorithms with considerably lower computational complexities, making them attractive in practice.

3 METHODS, ASSUMPTIONS, AND PROCEDURES

Robust Wideband Spectrum Knowledge Acquisition:

Let us consider a wide spectrum band of *B* Hz (where *B* can be in the order of hundreds of MHz to even a GHz) that is first segmented into several sub-bands [1]. Note that each sub-band may contain several channels possibly corresponding to different communications systems. As illustrated in Fig. 2, let us assume that there are N_b sub-bands. In general, scanning of these sub-bands spanning several non-contiguous frequency ranges can be achieved using reconfigurable antennas [15-23].

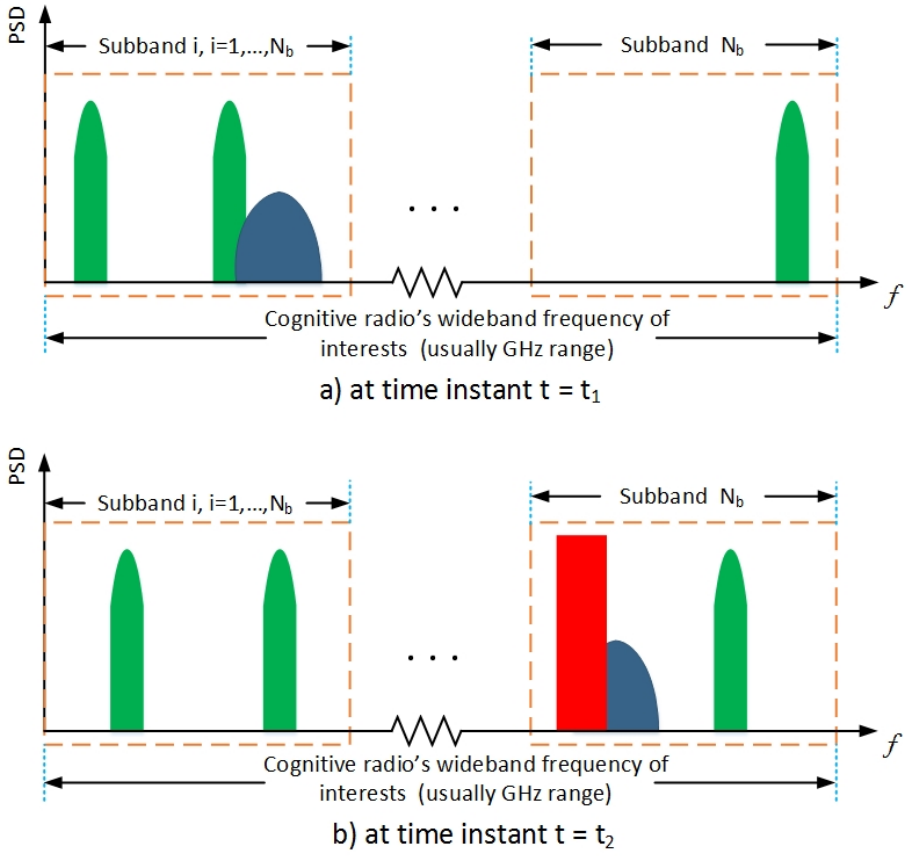


Figure 2. The spectrum of interest to a wideband cognitive radio is divided in to a set of N_b sub-bands. Each sub-band may contain possibly different types of signals and the amount of white-spaces can be time-varying.

Let us denote the N-length discrete-time sub-band signal by y where N is chosen to satisfy Nyquist sampling criteria. The discrete frequency domain representation y_f of this sub-band signal can then be written as [1,10,11]

$$y_f = Cy \tag{1}$$

where C is an N-point Discrete Fourier Transform (DFT) matrix.

When spectrum utilization within a sub-band is low, as in Fig. 1, we may expect y_f to be a sparse signal with respect to the frequency domain basis. While conventional Shannon-Nyquist sampling theory does not take in to account such sparsity of signals, compressive sampling allows this sparsity property of a signal to be exploited to detect spectral activities in each sub-band with a reduced number of samples. Indeed, suppose that the sensed signal within the sub-band of interest is sparse and that we only collect an M (where $M < N$) number of randomly selected samples from the signal y:

$$y_c = \Phi y = \Phi C^H y_f = Cy \tag{2}$$

where Φ is an M x N random sampling matrix and y_c is an M-length observation vector. As has been shown in [12,13], if the sampled signal is sparse then the signal y (or y_f) can indeed be reconstructed from the randomly compressive sampled (under-sampled) version y_c of y. In the presence of noise (2) becomes

$$y_c = \Phi y + w \tag{3}$$

where w is an M-length arbitrary noise vector.

Our previously proposed compressive sampling-based robust spectrum estimator allows us to obtain better spectral estimation in the presence of possibly non-Gaussian noise and interference [10]. This can improve spectral activity detection performance of the cognitive radio when indeed noise is non-Gaussian. Specifically, the proposed algorithm constructs the frequency-domain sub-band signal from y_c by solving the following optimization problem [11]:

$$y_f^* = \arg\min_{y_f \in C^N} l_H\left(y_c - Ay_f\right) + \gamma \left\| y_f \right\|_{l_1}. \tag{4}$$

where γ is a smoothing parameter that balances between the l_1-norm (sparsity) and the Huber cost function defined as

$$l_H(x) = \begin{cases} x^2/2 & if\ |x| \le \delta_H \\ \delta_H\left(|x| - \frac{\delta_H}{2}\right) & if\ |x| > \delta_H \end{cases}. \quad (5)$$

which is known to be optimal against ϵ – contaminated Gaussian noise [1,24,25].

Reinforcement Learning-based Sub-band Selection for Wideband Spectrum Sensing in Cognitive Radios:

Again, as shown in Fig. 2, we assume that the spectrum of interest to the WACR is divided into N_b sub-bands and that each sub-band may include a different number of communication channels. We denote by M_i the number of channels in the *i*th sub-band. Let $\mathcal{K} \in \{1,2,\dots\}$ denote the set of time slot indices. For simplicity, we assume the channel state to be constant within a single time slot. At any given time, each channel state can take two possibilities: Either occupied by another radio system (busy) or available to be used by the WACR (idle). We assume that this idle/busy state of each channel evolves according to a two-state (0/1 state), first order Markov chain [1,8,26].

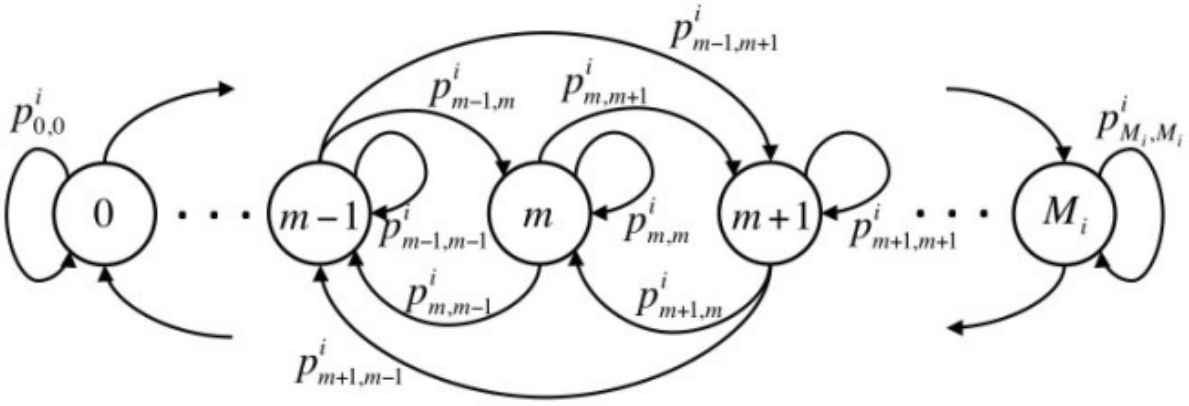


Figure 3. Markov chain model for the *i*-th sub-band when the state is defined to be the number of idle channels in the sub-band [1].

Sub-band selection decisions by a WACR will depend on its performance objective. In the following, we will assume that the goal of the WACR is to find the sub-band that has the largest number of idle channels, as proposed in [1]. Hence, we may define a new state $S_i[k]$ denoting the number of the idle channels in *i*th sub-band at time k, where $S_i[k] \in \{0,1,\dots,M_i\}$. If channel

idle/busy dynamics were to be Markov, as assumed above, then the dynamics of this new state $S_i[k]$ will also be Markov [1]. Figure 3 shows the sub-band Markov model, with $p_{s,s'}^{i}$ denoting the transition probability of the ith sub-band from state s to state s' [1]. The overall spectrum state at time k can then be defined as $\boldsymbol{S}[\boldsymbol{k}] = \{S_1[k], S_2[k], \dots, S_{N_b}[k]\}$. Let us denote by $\mathcal{S}$ the set of all the possible states $\boldsymbol{S}[\boldsymbol{k}]$ may take. Then, the set $\mathcal{S}$ has Z possible states, where $Z = \prod_{i=1}^{N_b}(M_i + 1)$.

At any given time, a WACR can only observe a single sub-band out of the total N_b sub-bands due to realistic hardware constraints. Hence, the sub-band selection problem can be considered as a decision making problem in which the system state can only be observed partially. Since we have assumed that the underlying system dynamics are Markov, thus the sub-band selection problem could be modeled as a partially observable Markov decision process (POMDP) [1]. We may define the selection process at time k as taking an action $a[k] \in A$ with the action space $A = \{1,2,\dots,N_b\}$ representing the set of sub-band indices. Let $Y[k]$ represent a partial observation corresponding to state $\boldsymbol{S}[\boldsymbol{k}]$ and $r(\boldsymbol{S}[\boldsymbol{k}], a[k])$ represents the immediate reward from taking action $a \in A$ when in state $\boldsymbol{S}[\boldsymbol{k}]$ at time k. We define the reward to be the number of idle channels available in the a-th sub-band at time $k+1$, if action a (i.e. the sub-band a) was chosen when in state $\boldsymbol{S}[\boldsymbol{k}]$ at time k [1,8,26]. Note that action $a[k]$ will be selected before observing $Y[k]$ corresponding to the current state at time k. Instead, what is available to the WACR is the history made of past observations, actions and the associated rewards up to the current time k denoted by $h[k]$.

Given all the available information up to time k, we may define a posteriori probability $b_m[k]$ as our belief that the current state $\boldsymbol{S}[\boldsymbol{k}]$ is s_m. The set of all a posteriori probabilities corresponding to all possible states is called the belief state vector $\boldsymbol{b}[k] = [b_1[k], b_2[k], \dots, b_Z[k]]^T$, with $b_m \in [0,1]$ for $m = 1, \dots, Z$ [1]. It is a well-known result that the belief state vector is a sufficient statistic for optimal decision making in a POMDP [1,27]. Thus, when making a decision, instead of taking into account all the history information $h[k]$, we may rely only on the belief state $\boldsymbol{b}[k]$.

Finding an optimal policy for the sub-band selection POMDP, however, leads to many challenges. First, it may require high computational complexity due to the continuous state space of the belief state vector [1,8,28-30]. Second, a policy needs to be computed in real-time. Moreover, we need the knowledge of sub-band Markov model parameters and, in particular, the transition probabilities of the model to be able to update the belief state vector. In addition, these model parameters may vary with time due to the dynamic nature of the wireless environment. These all make any attempt to directly compute an optimal policy complicated. As an alternative, we may use machine learning in which a WACR may attempt to learn an optimal policy instead of computing one [1]. A particular type of machine learning approach, called reinforcement learning, could especially be suited when underlying state dynamics are Markov as assumed in our model [1,14].

Q-learning is one of the most widely used reinforcement learning approaches [1,14,31], in which the algorithm maintains a Q-table containing Q-values denoted by $Q(S,a)$ that represents a measure of goodness resulting from taking an action a (selecting the a-th sub-band) when in state S. Hence, if the selected sub-band contains a large number of idle channels this may lead to a high reward and, consequently, a high Q-value. However, Q-learning is not directly applicable to our POMDP sub-band selection problem. Thus, in our work we have resorted to an extension of Q-learning called the replicated Q-learning [1,32].

The replicated Q-learning algorithm attempts to reinforce the actions that lead to better outcomes from a given state. Each time an action is selected in a given state, the Q-table is updated as in [1,26]

$$Q(s_m, a[k-1]) \leftarrow Q(s_m, a[k-1]) + \alpha b_m[k-1]\left[r(s_m, a[k-1]) + \gamma \max_a Q(b[k], a) - Q(s_m, a[k-1])\right] \quad (6)$$

Recall that, the reward $r(s_m, a[k])$ represents the number of idle channels available in the selected sub-band and b_m is the m-th element of the belief state vector $\boldsymbol{b}$. We denote by $\alpha \in (0,1)$ the learning rate while the parameter $\gamma \in [0,1)$ represents the discount factor. Future actions (sub-band selections) will then be selected based on the updated Q-values:

$$a^* = \arg\max_{a \in A} Q(b[k], a) \quad (7)$$

where $Q(b[k],a)$ is the average of the Q-values when taking action a from all possible states given the belief state $\boldsymbol{b}$, given by $Q(b[k],a)=\sum_m b_m Q(s_m,a)$.

As with any adaptive or learning algorithm, Q-learning may also get trapped at a local optimal leading to a policy that may not be the optimal. In order to avoid this problem we may define a new parameter called exploration rate $\epsilon \in (0,1)$. Depending on the exploration rate, the CR can switch between selecting the action characterized by (7) or just randomly selecting an action out of all possible actions [1]:

$$a^* = \begin{cases} \arg\max_{a\in A} Q(b[k],a) & \textit{with probability } 1-\varepsilon \\ \sim U(A) & \textit{with probability } \varepsilon \end{cases} \tag{8}$$

where $U(A)$ denotes the uniform distribution over the action set A. Choosing a high exploration rate may help in updating the entire Q-table and avoid being trapped in a sub-optimal policy. On the other hand, a low exploration rate will help in exploiting the already learned optimal actions. Thus, obtaining an optimal policy requires the selection of an appropriate exploration rate that could balance between the exploration and exploitation [1,8,14,26].

A New Spectrum State Definition for Interference Avoiding and Anti-jamming WACRs:

A common situation in which cognitive communications can be a great asset is when reliable communications is needed in the presence of either unintentional interference and/or deliberate jammers. In this case, each WACR will attempt to avoid the jammer signals as well as the other WACRs' transmission. In these situations, we may reduce the computational complexity of sub-band selection algorithms by defining a binary-valued state for each sub-band: Either the sub-band is free of interference and jammers according to a certain criterion (state 1) or it is not (state 0) [2].

Under certain conditions, it can be argued that this state can reasonably be modeled as being Markov. In our current work, we are developing these justifications and future work will employ this new state definition to design lower complexity sub-band selection and other cognitive algorithms for wideband autonomous cognitive radios.

4 RESULTS AND DISCUSSION

Robust Wideband Spectrum Knowledge Acquisition:

Previously we had compared the performance of our compressive sampling-based robust spectrum estimate of the sub-band sensed signal with that of the Gaussian-optimal periodogram estimate. In order to evaluate this performance in a simplified simulation scenario we considered a sub-band signal y of length N = 128 composed of three active signals x_1, x_2 and x_3 located at center frequencies (discrete) of 5, 20 and 40, respectively, as shown in the top row of Fig. 4. The corresponding signal amplitudes were arbitrarily chosen to be 8, 15 and 22. Each signal has a bandwidth of B = 7 (in discrete frequency) around its center frequency. For random compressive sampling, the sensing matrix was drawn according to a normal distribution.

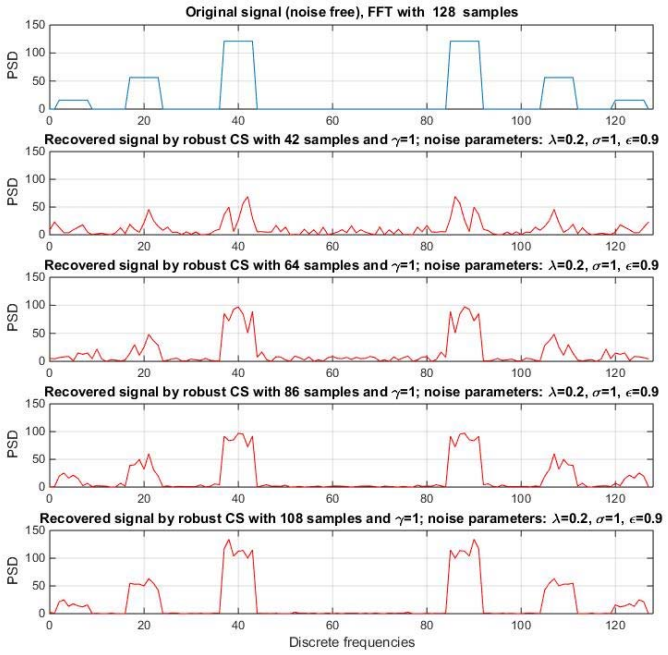


Figure 4. Power spectral densities (PSD) of original and recovered signals by robust compressive sampling with varying compression ratios 33%, 50%, 67%, 84% (top to bottom).

For completeness, Fig. 4 shows the reconstructed sub-band signal by solving (5) as we vary the compression ratio of the number of samples from 33% to 84% (with respect to the required Nyquist rate) in the presence of Gaussian-Laplacian mixture noise with ϵ = 0:9. Clearly, even at

the high compression ratio of 67%, the performance of the reconstructed signal seems to be reasonable enough for signal activity detection in the presence of noise. Moreover, as we show in Fig. 5 below, previously we also observed that our proposed algorithm with a reduced number of (compressed) samples can provide almost comparable or better performance compared to the traditional periodogram estimate.

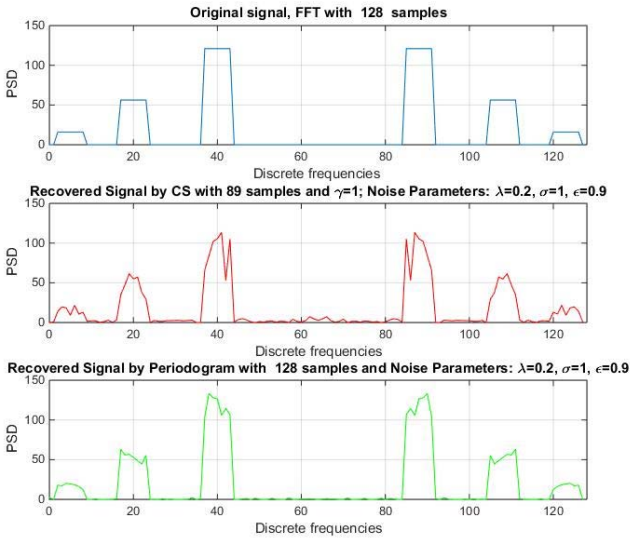


Figure 5. Power spectra of original signal and those recovered by the robust compressive sampling algorithm (with 89 samples) and by the Periodogram (with 128 samples).

During the last performance period, we continued with this performance analysis to obtain more detailed performance characteristics of the proposed compressive sensing-based robust spectral estimator. Figure 6 shows the robustness of the proposed approach even with fewer samples (89 samples in this case) against non-Gaussian noise as the amount of non-Gaussianity increases, where we used the performance metric to be the normalized root mean-squared error defined as below [10]:

$$NRMSE_{PSD} = \frac{\left\| y_f^2 - y_f^{*2} \right\|_{l_2}}{\left\| y_f^2 \right\|_{l_2}}. \tag{9}$$

Figure 7 shows how normalized root mean-squared error (NRMSE) of the power spectrum estimate decreases with both number of (compressive) samples used as well as with increasing signal to noise ratio (SNR).

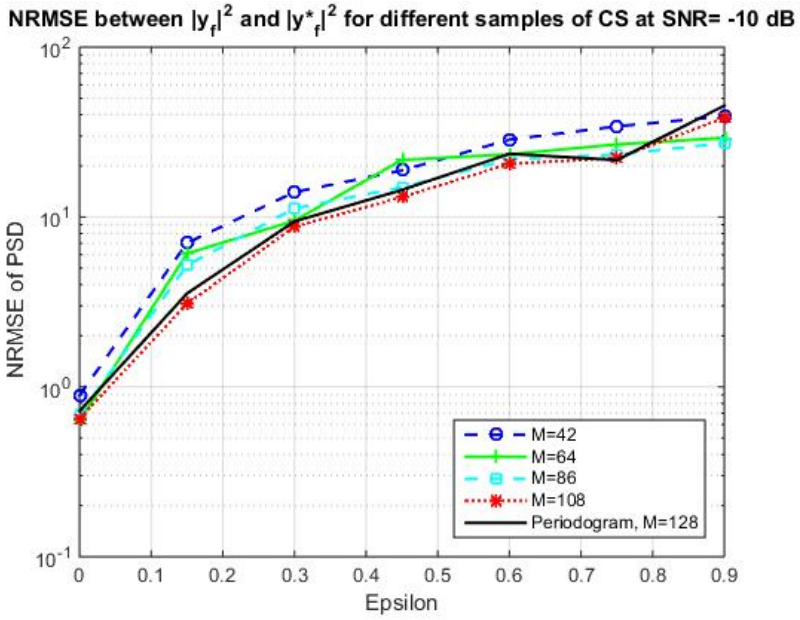
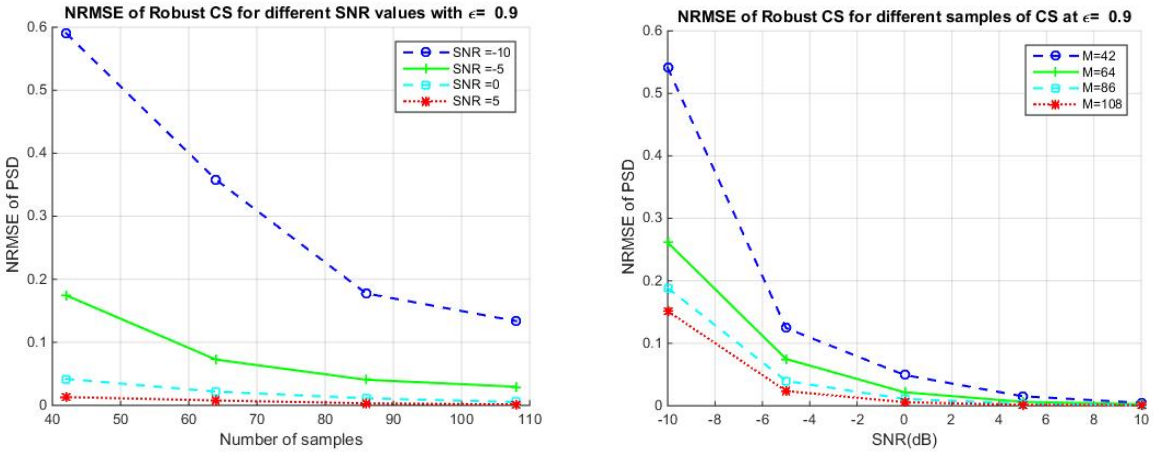


Figure 6. $NRMSE_{PSD}$ between original and recovered signals by robust compressive sampling (for different number of samples) and Periodogram (128 samples).



(a) **(b)**

Figure 7. Normalized root mean-squared error performance of the compressive sampling based robust spectrum estimator. (a) The $NRMSE_{PSD}$ between original and recovered signals by robust compressive sampling for different number of samples. (b) The $NRMSE_{PSD}$ between original and recovered signals by robust compressive sampling at different SNRs with different number of samples.

In our simulation set-up of a wideband spectrum knowledge acquisition framework, the wide spectrum of interest to the WACR is defined as composed of a set of non-contiguous frequency ranges and divided in to a set of N_b sub-bands [1,8,26]. The model allows these sub-bands to have different number of signals with different properties and be either ON or OFF at any given time. The receiver processing modules perform the down-conversion, analog-to-digital conversion of input/output Q channels and the subsequent cognitive processing in the digital domain, as proposed in [1]. As a baseline comparison scenario, we have implemented spectral activity detection and feature extraction based on the periodogram approach. This simulation was used to analyze the trade-off in detection performance and feature accuracy (in this case bandwidth) as a function of the detection threshold and the smoothing window length.

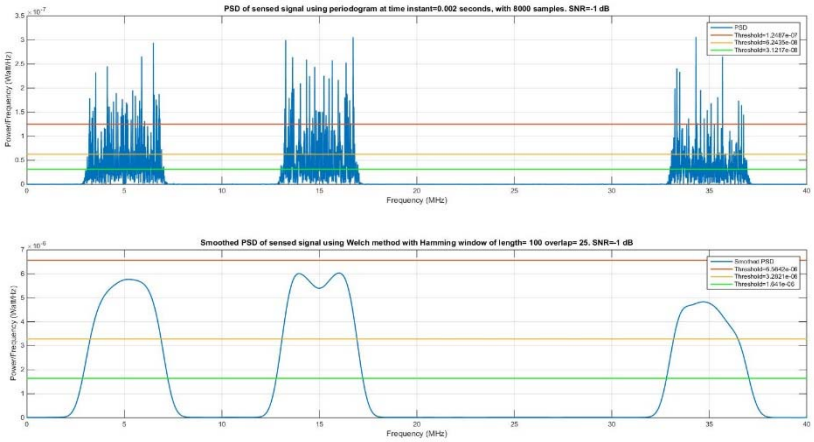


Figure 8. Dependence of detection probability and bandwidth estimation error on smoothing and Neyman-Pearson threshold (SNR -1 dB).

Figure 8 shows the direct periodogram estimated from the sensed sub-band signal and the smoothed periodogram estimate (after down-converting to the baseband). In this case, there are 3 active signals in this sub-band of width 40MHz (Note that the universal software radio peripheral (USRP) 2943R SDR board from National Instrument has an instantaneous bandwidth of 40MHz). Shown also on Fig. 8 is the exact Nayman-Pearson (NP) threshold computed from theory as well as two other possible empirical threshold values. Note that, the NP threshold is computed to maximize the detection probability of signals. However, it is observed that in many cases of our assumed RF environment this threshold may lead to larger mean squared errors in the estimated

bandwidth of the detected signal. In order to obtain a compromise between the detection probability and the bandwidth estimation accuracy, it has been observed that we need a trade-off between the smoothing window length and the detection threshold.

TABLE 1: Bandwidth estimation error of smoothed signal with 8000 samples. (run time: 50 trials).

SNR (dB)

Threshold		**-5dB**	**-1dB**	**1dB**
	NP threshold	No signal detected	13.5003	1.6786
	0.5 x NP threshold	No signal detected	0.0896	0.0811
	0.25 x NP threshold	1.6724	0.0144	0.0712

Table 1 shows the bandwidth estimation errors achieved by using different threshold values on the smoothed periodogram estimate. The results are averaged over 50 trials. In each trial the sub-band spectrum is estimated using 8000 signal samples. The corresponding detection probabilities for the case of SNR=-1dB are shown in Table 2. As can be seen for lower threshold values (0.5xNP threshold and 0.25xNP threshold), the detection performance is perfect whereas the theoretical NP threshold results in a detection probability of only 0.32. This is because it is optimal for the non-smoothed periodogram and not for the smoothed periodogram. Table 1 shows that these lower thresholds also lead to much smaller bandwidth estimation errors compared to that with the exact NP threshold.

TABLE 2: Probability of detection of smoothed signal with 8000 samples. (run time: 50 trials).

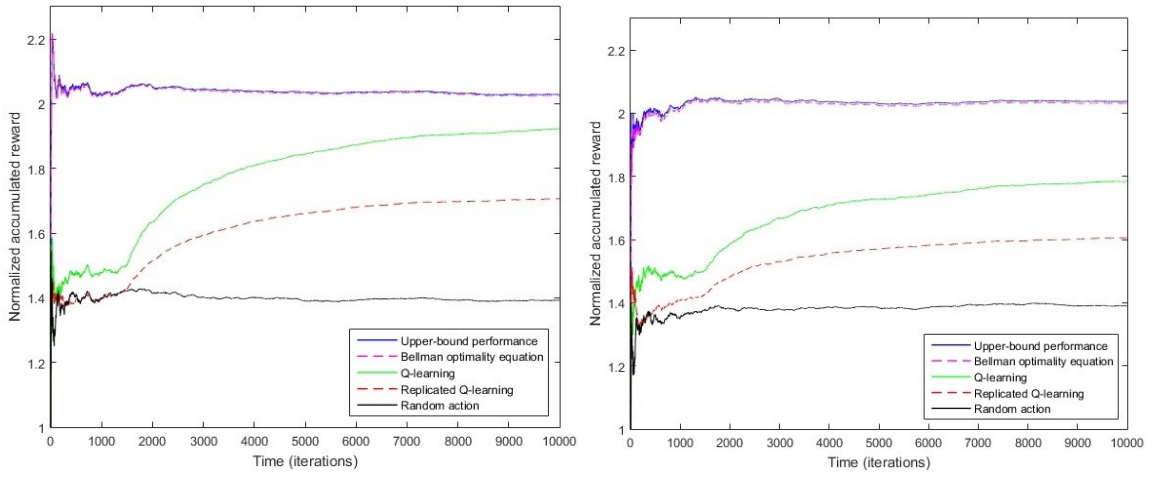
SNR (dB)

Threshold		**-1dB**
	NP threshold	32%
	0.5 x NP threshold	100%
	0.25 x NP threshold	100%

Reinforcement Learning-based Sub-band Selection for Wideband Spectrum Sensing in Cognitive Radios:

The performance of our proposed replicated Q-learning algorithm was compared against four benchmarks. First is called the upper-bound performance. It was obtained by assuming that the WACR may observe the exact state at time k before selecting action $a[k]$. Second is the performance of the optimal sub-band selection policy obtained by solving the Bellman-optimality equation [1,33]. This, in other words, is the optimal performance of the associated Markov decision process (MDP) problem. Third, we used a Q-learning algorithm under the assumption that the states are completely observable. Fourth, and finally, we used the performance of a random sub-band selection scheme in which all sub-bands are selected with equal probabilities.

In the spectrum model, we assumed that there are N_b = 3 sub-bands. The total number of channels in the spectrum is 8 channels in which the second sub-band contains 2 channels and the remaining 6 channels are divided equally in the first and the third sub-bands [26]. All channels are assumed to have the same bandwidth. In addition, the dynamics of these channels are assumed to be independent of each other. Each simulation was carried out over 10,000 iterations. We observed that about 1,500 iterations were needed for the Q-table to be considered as converged.



(a) **(b)**

Figure 9. Comparison of normalized accumulated reward of sub-band selection policies. (a) Relatively low exploration after convergence with $\epsilon = 0.01$ (b) Relatively high exploration after convergence with $\epsilon = 0.3$.

Figure 9 compares the performance of the replicated Q-learning with the other four methods mentioned above [26]. As our performance metric, we used the normalized accumulated reward, defined as

$$R_N = \frac{1}{N}\sum_{k=1}^{N} r(S[k], a[k]) \tag{10}$$

where N is the number of iterations. Unless noted otherwise, a discount factor of $\gamma = 0.2$ was used. In addition, initially we allowed a high exploration rate of $\epsilon = 0.8$ and a learning rate of $\alpha = 0.4$. After convergence, we reduced the learning rate and the exploration rate to $\alpha = 0.1$ and ϵ =0.01, respectively.

As can be seen from Fig. 9(a), the random sub-band selection policy can only achieve about a 68% of that of the optimal policy. As one would expect, the performance of both Q-learning and replicated Q-learning lie somewhere between the optimal and random-action policies. It can be seen from Fig. 9(a) that Q-learning converges about 95% of the performance achieved by the optimal policy. On the other hand, the replicated Q-learning algorithm achieves about 84% of the performance of the optimal policy. This is significant in three ways: First, it shows that the replicated Q-learning can indeed provide noticeably better performance than simply selecting random sub-bands for sensing. Second, its performance is not that far from that of the optimal sub-band selection policy that requires complete state observability. Third, and final, is the fact that replicated Q-learning achieves about 88% of the performance of the Q-learning which is a better performance upper-bound for comparison [26].

Recall that the choice of ϵ is a trade-off between the exploration and exploitation. Figure 9(b) shows the effect of using a relatively larger value of $\epsilon = 0.3$ after the convergence compared to Fig. 9(a). As can be seen from Fig. 9(b), performance of both Q-learning and replicated Q-learning has degraded. The Q-learning achieves 88% of the optimal performance, while replicated Q-learning achieves only about 79% of the optimal performance. The reason is that the higher exploration rate leads to too much exploration. The WACR selects random actions more often than in Fig. 9(a) as opposed to exploiting the already learned better actions.

5 CONCLUSIONS

During the just finished performance period, we developed a compressive sampling-based robust wideband spectrum knowledge acquisition framework suitable for a wideband autonomous cognitive radio (WACR). The proposed method augments the Huber cost function with an additional l_1-norm penalty term in order to find a sparse spectrum estimate while achieving robustness against possibly non-Gaussian noise. We observed that the proposed approach can improve the wideband spectrum sensing performance in two important ways: 1) the required number of samples can be reduced and 2) the estimation performance can be better than that of the conventional periodogram. Motivated by the fact that spectrum knowledge acquisition involves both detection and identification of spectral activity and that the signal classification can more efficiently be achieved based on features, we performed extensive experiments to gain an understanding of the heuristics involved in properly choosing the smoothing window and the detection threshold to achieve a trade-off between the detection probability and the bandwidth estimation errors.

In order to develop a sub-band selection mechanism suitable for a WACR, we modeled the sub-band selection problem as a partially observable Markov decision process (POMDP), in which only a single sub-band can be sensed at any given time out of all available sub-bands in the spectrum of interest. This model was then used to develop an effective, low-complexity policy to select the sub-bands based on a machine learning algorithm called the replicated Q-learning. Simulation results showed that the proposed replicated Q-learning approach can provide a substantial improvement over the random sub-band selection policy. We also showed that it is better in practice to use a relatively larger exploration rate at the beginning so that fast learning can be achieved. However, after the convergence the exploration rate shall be reduced accordingly to reap the benefits of the already learned actions.

Finally, we observed that the original approach proposed in [1] to define the spectrum sub-band state may lead to sub-band selection algorithms with unacceptably high computational complexities. As a result, we have been investigating new mathematical definitions for the sub-band state that may specifically be applicable for certain contexts but can lead to decision policies, learning algorithms and cognitive protocols with sufficiently low computational complexities.

6 RECOMMENDATIONS

Spectrum is a battlefield. As more nations develop advanced communications and radar technologies, mission success will be challenged by various adverse spectrum conditions including both deliberate as well as inadvertent interference. Moreover, encroachment on previously-allocated spectrum resources by the commercial and unlicensed/unauthorized users can only expected to grow in the coming years. The combination of these traditional as well as evolving spectrum challenges requires future communications technologies to be intelligent, self-aware and spectrally agile. Unlike cognitive radios treated in most of the existing literature, the WACRs proposed in [1] and pursued in this project are radios with these defining characteristics that will allow autonomous radio communications over non-contiguous wide spectrum bands in the presence of adverse conditions. It is the WACR technology that has the potential to revolutionize the future communications systems. Thus it is recommended that future efforts be focused on fully developing WACR technology.

The compressive sampling-based robust wideband spectrum sensing approach developed in this project can form the first step in the spectrum knowledge acquisition framework of a WACR. The approach can help handle large sub-band bandwidths with realistic hardware constraints and be robust against non-Gaussian noise statistics. With the experience gained from developing our reinforcement learning aided sub-band selection algorithm for a WACR, we believe that the new state definition can lead to algorithms with considerably lower computational complexities. Future efforts must thus be focused on reformulating some of the models we had developed in the past using this new state definition and developing machine learning-based cognitive communications algorithms suitable for wideband autonomous cognitive radios.

REFERENCES

1. S. K. Jayaweera, Signal Processing for Cognitive Radios, 1st ed. New York, NY, USA: John Wiley & Sons Inc., 2014.

2. S. Machuzak and S. K. Jayaweera, "Reinforcement learning based anti-jamming with wideband autonomous cognitive radios," *IEEE/CIC International Conference on Communications in China (ICCC)*, Chengdu, China, July 2016.

3. J. Mitola and G. Q. Maguire Jr, "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, 1999.

4. S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.

5. S. K. Jayaweera and C. G. Christodoulou, "Radiobots: Architecture, Algorithms and Realtime Reconfigurable Antenna Designs for Autonomous, Self-learning Future Cognitive Radios," University of New Mexico Technical Report, EECE-TR-11-0001, Albuquerque, NM, Mar. 2011.

6. S. K. Jayaweera, Y. Li, M. Bkassiny, C. G. Christodoulou and K. A. Avery, "Radiobots: The autonomous, self-learning future cognitive radios," *IEEE Intelligent Sig. Proc. and Commun. Systems (ISPACS'2011)*, Chiangmai, Thailand, December 2011.

7. M. Bkassiny, S. K. Jayaweera, Y. Li, and K. A. Avery, "Wideband spectrum sensing and non-parametric signal classification for autonomous self-learning cognitive radios," *IEEE Transactions on Wireless Commun.*, vol. 11, no. 7, pp. 2596–2606, July 2012.

8. Y. Li, S. K. Jayaweera, M. Bkassiny and C. Ghosh, "Learning-aided Sub-band Selection Algorithms for Spectrum Sensing in Wide-band Cognitive Radios," *IEEE Trans. in Wireless Commun.*, vol. 13, no. 4, pp. 2012 – 2024, April 2014.

9. M. Bkassiny, S. K. Jayaweera, and Y. Li, "Multidimensional Dirichlet Process-based Non-Parametric Signal Classification for Autonomous Self-Learning Cognitive Radios," *IEEE Trans. in Wireless Commun*., vol. 12, no. 11, pp. 5413-5423, November 2013.

10. P. Das and S. K. Jayaweera, "Robust wideband spectrum sensing with compressive sampling in cognitive radios," *IEEE Vehicular Technology Conference (VTC Fall)*, Boston, MA, September 2015.

11. Z. Tian and G. B. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, April 2007.

12. D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

13. E. J. Candes and M. B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

14. R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, MIT Press, Cambridge, MA, 1998.

15. Y. Tawk, J. Costantine, K. Avery, and C. G. Christodoulou, “Implementation of a cognitive radio front-end using rotatable controlled reconfigurable antenna,” *IEEE Trans. on Antennas and Propagation*, vol. 59, no. 5, pp. 1773-1778, May 2011.

16. C. G. Christodoulou, Y. Tawk, S. Lane, and S. Erwin, “Reconfigurable antennas for wireless and space applications,” *Proceedings of the IEEE*, vol. 100, no. 7, pp. 2250-2261, July 2012.

17. Y. Tawk, M. Bkassiny, G. El-Howayek, S. K. Jayaweera, K. Avery and C. G. Christodoulou, “Reconfigurable front-end antennas for cognitive radio applications,” *IET Microwaves, Antennas and Propagation*, vol. 5, no. 8, pp. 985-992, June 2011.

18. E. Ebrahimi, J. R. Kelly, and P. S. Hall, "Integrated wide-narrowband antenna for multi-standard radio," *IEEE Trans. on Antennas and Propagation*, vol. 59, no. 7, pp. 2628-2635, July 2011.

19. G. T. Wu, et al., “Switchable quad-band antennas for cognitive radio base station applications,” *IEEE Trans. on Antennas and Propagation*, vol. 58, no. 5, pp. 1468-1476, May 2010.

20. T. Aboufoul, et al.,“Reconfiguring UWB monopole antenna for cognitive radio applications using GaAs FET Switches,” *IEEE Antennas and Wireless Propagation Letters*, vol. 11, pp. 392-394, 2012.

21. Y. Tawk, J. Costantine, and C. G. Christodoulou, “A varactor based reconfigurable filtenna,” *IEEE Antennas and Wireless Propagation Letters*, vol. 11, pp. 716-719, 2012.

22. Y. Tawk, and C. G. Christodoulou, “A new reconfigurable antenna design for cognitive radio applications,” *IEEE Antennas and Wireless Propagation Letters*, vol. 8, pp. 1378-1381, December 2009.

23. Y. Tawk, S. K. Jayaweera, C. Christodoulou, and J. Costantine, “A comparison between different cognitive radio antenna systems,” in *Proc. Int. Symp. ISPACS*, Chiangmai, Thailand, December 2011.

24. M. Bkassiny and S. K. Jayaweera, “Robust, non-Gaussian wideband spectrum sensing in cognitive radios,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 11, pp. 6410–6421, 2014.

25. P. J. Huber and E. M. Ronchetti, Robust Statistics., 2nd ed. New York, NY, USA: John Wiley & Sons Inc., 2009.

26. M. A. Aref, S. Machuzak, S. K. Jayaweera and S. Lane, "Replicated Q-learning based sub-band selection for wideband spectrum sensing in cognitive radios," *IEEE/CIC International Conference on Communications in China (ICCC)*, Chengdu, China, July 2016.

27. C. Striebel, “Sufficient statistics in the control of stochastic systems,” *J. Math. Anal. Appl.*, vol. 12, pp. 576–592, 1965.

28. G. E. Monahan, “A survey of partially observable Markov decision processes theory, models and algorithms,” *Management Science*, vol. 28, no. 1, pp. 1–16, January 1982.

29. E. J. Sondik, “The optimal control of partially observable markov processes,” Ph.D. dissertation, Stanford University, 1971, https://searchworks.stanford.edu/view/2231661, n.d.

30. R. D. Smallwood and E. J. Sondik, “The optimal control of partially observable Markov processes over a finite horizon,” *Operations Research*, vol. 21, pp. 1071–1088, 1973.

31. C. Watkins and P. Dayan, “Q-learning,” *Machine Learning*, vol. 8, no. 3, pp. 279–292, 1992.

32. M. L. Littman, A. R. Cassandra, and L. P. Kaelbling, “Learning policies for partially observable environments: Scaling up,” in *Proc. 12th Int. Conf. Mach. Learn.*, Tahoe, City, CA, 1995.

33. D. P. Bertsekas, Dynamic Programming and Optimal Control, 3rd ed. Belmont, MA: Athena Scientific, 2005, vol. 1.

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

DFT	Discrete Fourier Transform
DSS	Dynamic Spectrum Sharing
EMI	Electromagnetic Interference
MDP	Markov Decision Process
NRMSE	Normalized Root Mean-Squared Error
NP	Neyman-Pearson
POMDP	Partially Observable Markov Decision Process
PSD	Power Spectral Density
RF	Radio frequency
SDR	Software-defined Radio
SNF	Signal to noise ratio
USRP	Universal Software Radio Peripheral
WACR	Wideband Autonomous Cognitive Radio

DISTRIBUTION LIST

DTIC/OCP
8725 John J. Kingman Rd, Suite 0944
Ft Belvoir, VA 22060-6218 1 cy

AFRL/RVIL
Kirtland AFB, NM 87117-5776 2 cys

Official Record Copy
AFRL/RVSW/Khanh Pham 1 cy